IJCSC 0973-7391 Volume 8 • Issue 2 March 2017 - Sept 2017 pp. 174-180 available online at www.csjournals.com

A Review on Various Load balancing Algorithm for Cloud Computing

Nutan[1], Dr. Sanjeev Kumar[2]
[1]M.Tech. Scholar, [2]Assistant Professor
Department of CSE, GJUS&T, Hisar, Haryana, India.
nutanjatain75@gmail.com, sanjukhambra@yahoo.co.in

Abstract: Cloud computing is a technical development that focuses on the way to design computing systems, develop applications, and leverage existing services for building software. Load balancing is a critical issue for cloud computing. Load balancing is a technique to allocate work between two or more computers, networks, processors or memory devices in order to channelize the resources in an efficient manner and to get optimized response times and throughputs. In this paper firstly load balancing is discussed in detail, and then review of previous work is provided. After that some load balancing algorithms are studied and compared.
Keywords: ESCE, VM, TLB, RR.

1. Introduction

Cloud Computing is the new style of computing, it provides different type of services like- servers, storage, application, database, networking, software and more over the internet on the user's computer or devices. It is a kind of computing that spotlights on sharing computing resources rather of having the local host or individual gadgets to oversee applications. Definition according to the NIST (official), "cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. Cloud computing is a term for anything that involves delivering hosted services over the Internet. The main three types of categories are there in these services: IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service) and SaaS (Software-as-a-Service) [2].Cloud computing has received more and more attention and been adopted in both the industry and academics because of its vital characteristics such as elasticity, quality of service, guaranteed and on-demand resource provisioning model. Cloud computing allows renting infrastructure, services and runtime environments on a pay- per-use basis. This principle finds several practical applications and then gives different images of cloud computing to different people. Cloud computing also provides an opportunity for integrating additional capacity or new features into existing systems. For cloud providers, the suitable load balancing [3] mechanisms are important to them, which can reduce the load of overload nodes and improve resource utilization of under-loading nodes. Load balancing is an important term in cloud computing environments and generally distributed systems which affect the system performance dependent on the amount of work allocated to the system for a specific time period. The process of redistributing the general system workload among system resources for improving resource utilization and system performance is called load balancing [4]. Load balancing play crucial role in every virtual machine in the cloud system bears the same amount of workload. Through this, throughput can be increased and the response time gets decreased, and hence high users' satisfactions will be achieved. In cloud computing environment, all the available resources are shared but if they are not properly managed and distributed then result will be resource wastage. Today the needs of resources are increasing rapidly. Therefore, it is crucial to allocate the resources rightly so we are moving towards dynamic resource allocation. For this we are making use of virtualization technique that migrate virtual machines to physical machines effectively. Virtualization is a technique which allows complete installation of one machine on to the other one. The load balancing techniques work like to select next server node and to transfer new incoming requests to that selected node. It is a process to lift the whole workload to the individual nodes which are unused or not having much loads in order to get efficient utilization of resources & to improve the response time of some job also at the same time avoiding the situation where a few nodes are loaded heavily and rest of nodes are slightly loaded.

The organization of the paper is as follows: In section II, load balancing is discussed. In section III, review of some related work is done. In section IV, some load balancing algorithms are studied. In section V and VI, some metrics for load balancing algorithms are given and algorithms are compared on the basis of these metrics. Section VII, concludes the paper.

2. Load balancing

Load balancing is defined as an approach to increase and improve the performance of two or more nodes or links connected nodes by the redistribution or the reassignment of load. Load balancing is used to distribute a larger processing load to smaller processing nodes for enhancing the overall performance of system [7]. In cloud

computing atmosphere load balancing is required to distribute the dynamic local workload uniformly between all the nodes.

- Load balancing helps in fair allocation of computing resource to achieve a high User satisfaction and proper Resource utilization.
- Load balancing is a technique that helped networks and resources by providing a Maximum throughput with minimum response time.
- It is very important to estimate proper load, need to do comparison of all load, stability of all different systems, performance of purposed system, interaction between all the nodes and nature of work to be transferred while developing a load balancing algorithm. The most important thing is selecting the nodes and it also includes many other ones. CPU load, amount of memory required combine together to calculate the load of machine.

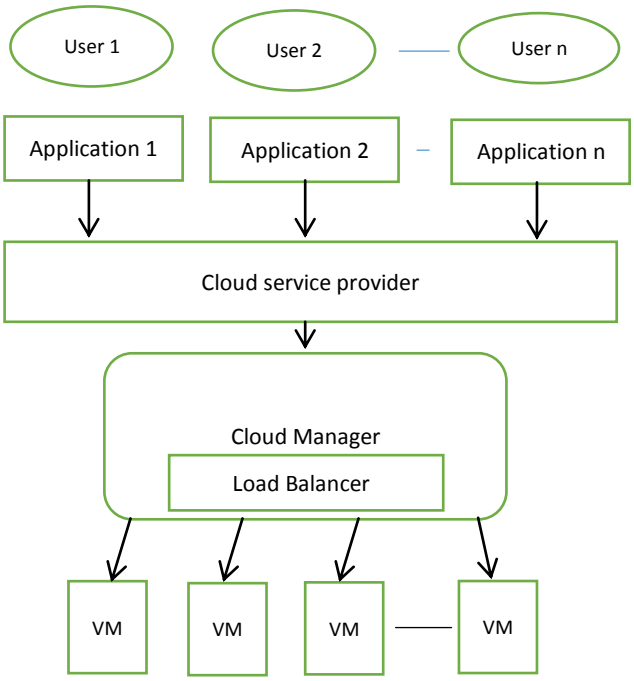

Figure1: Load balancing

In figure 1, load balancing process is explained. Users interact with cloud service providers through applications. Load balancer, which is in direct control of cloud manager, plays the main role of balancing the workload over the entire system. Load balancer allocates and de-allocates virtual machines to the nodes. Virtual machines perform certain tasks.

Load balancing Algorithm

Load balancing Algorithms are mainly divided into two categories as shown in figure 2:

- Static load balancing algorithm
- Dynamic load balancing algorithm

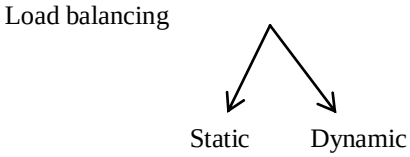

Figure 2: Types of Load Balancing Classification

Static approach:-

Algorithms which fall under this category need previous knowledge of the system and do not depend on the current state of the system [10]. These algorithms are of not pre-emptive nature. These aim to minimize the

execution time, communication overhead and delays. Static algorithm has a drawback that the task cannot be shifted at the time of execution to any other node to balance the load.

Dynamic approach:-

This method deliberated the current state of the system during load balancing decisions. No prior knowledge of the system is needed. Advantage of this approach is that it improves the overall performance of the system by shifting the load dynamically during execution time because its decision for balancing the load is based on the current state of the system. This approach is more suitable for widely distributed systems such as cloud computing.

Dynamic load balancing approaches have two types. They are distributed approach and non-distributed (centralized) approach. It is defined as following:

a) Centralized approach: - In centralized approach, only a single node is responsible for managing and distribution within the whole system. Other all nodes are not responsible for this [6].

b) Distributed approach: - In distributed approach, each node independently builds its own load vector. Vector collects information of other nodes load. All decisions are completed locally using local load vectors. Distributed approach is further suitable for widely distributed systems such as cloud computing [7].

3. Related works

There have been some works in the field of cloud computing and load balancing. In this section, study of various papers is done which are related to cloud computing and its chief concern of load balancing.

Shyam Patidar et. al. discussed architecture, components and opportunities of cloud computing. Authors have also reviewed some challenges of cloud, and classification of clouds also [8]. Author in [9] reviewed several cloud deployment and service models and presented the benefits of cloud computing over traditional IT service environment- including scalability, flexibility, reduced capital and higher resource utilization -are considered as adoption reasons for cloud computing environment.

Asha M L et. al. presented a study of round robin algorithm. Authors also provided implementation, and performance analysis of the algorithm based on different parameters [12].

Ray, Soumya et. al. discussed some merits and demerits of the cloud computing, major thrust is given on the study of load balancing algorithm and performance analysis on existing VM load balancing algorithm is done [13]. In the paper, the simulation setup has done for implementing three policies namely; Round Robin, Central queuing and Randomized in combination with various parameters.

Khiyaita et. al. addressed the state-of-the-art of load balancing in cloud environment [15]. Authors provided different classifications of load balancing, gave examples of some load balancing systems and discussed load balancing challenges in cloud computing.

Isam Azawi Mohialdeen et. al. discussed the various scheduling policies and performed comparative study among scheduling algorithms in cloud computing and addressed their need in cloud computing[17].

Ram Prasad Padhy et. al. described important notions of cloud computing, services provided by cloud providers and operations provided by the cloud [20]. Authors proposed a better load balance model for the public cloud which is based on the cloud partitioning concept with a switch mechanism that can choose different strategies in different situations

In [21] Round Robin and Throttled virtual machine load balancing policies have been compared. These load balancing techniques are considered in combination with optimized response time service policy and simulation is done by varying parameters to analyse the performance. Simulation results show that combination of throttled and optimized response time service broker policy performs better than round robin policy in heterogeneous cloud environments.

In this paper [22] some popular load balancing techniques have been taken into consideration, such as Round Robin, Throttled, Execution Load and First Come First Serve. Performance analysis of these algorithms is done considering parameters like average response time, average datacenter request servicing time and total cost. The simulation results in CloudAnalyst simulator show that round robin is the best performing algorithm.

Authors in [23] discussed about functioning of GUI based tool called as Cloud Analyst which can be used for studying the behaviour of huge scaled internet applications.

4. Load balancing algorithms

➢ Round Robin Algorithm

This algorithm is simpler then all available algorithms for load balancing and hence do not require complex algorithm implementations [10]. Architecture of Round robin Algorithm is shown in figure 3. It simply maintains a queue of incoming requests and allocates them VM in Time scheduling manner. Thus each request is allowed to be executed for specific time quantum only then after if it is still incomplete, it has to wait for its next round and if the request is complete it allows other process to take charge of that VM based on the algorithm. It attempts to

maintain equal workloads on all the available VMs. Even though there is equal workload distributions between processors but processing time for different processes are not identical. So at any some of the nodes may be heavily loaded and others stay idle [18].

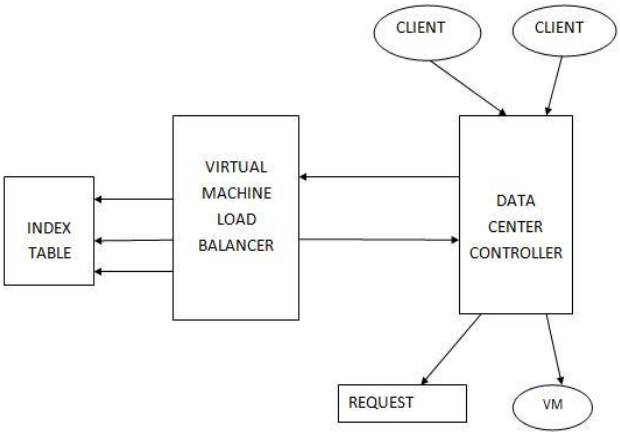

Figure 3: Architecture of Round robin Algorithm

➢ **Equally Spread Current Execution Algorithm (ESCE)**

In spread spectrum technique load balancer makes effort to preserve equal load to all the virtual machines connected with the data center [12]. ESCE algorithm is depicted in figure 4. Load balancer sustains an index table of VM as well as number of requests currently assigned to the Virtual Machine (VM).

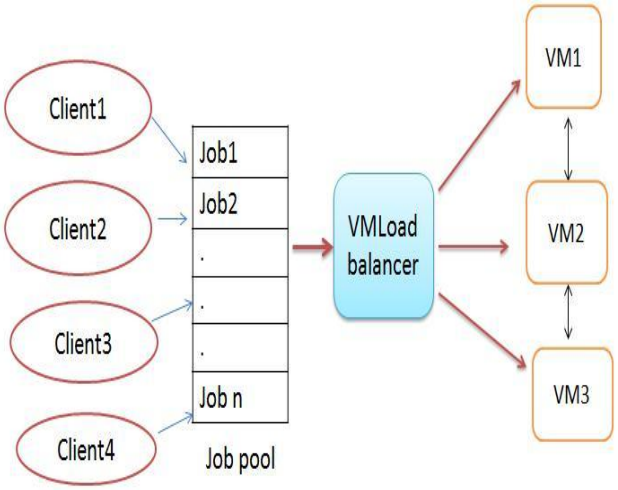

Figure 4: ESCE Algorithm

If the request comes from the data center to assign the new VM, it scans the index table for least loaded VM. If there is more than one VM is found than first recognized VM is selected for handling the request of the client/node, the load balancer also returns the VM id to the data center controller. [16] The balancer attempts to improve the response time and processing time of a job. A shortcoming of this algorithm is that it is not fault tolerant and problem of single point of failure occurs [19].

➢ **Throttled Load Balancing Algorithm (TLB)**

Throttled algorithm is shown in figure 5; in this algorithm the load balancer maintain index table of virtual machines and their states (busy/idle).The client/server first makes a request to data center to find a suitable virtual machine (VM) to perform the recommended job [14]. The data center queries the load balancer for allotment of the VM. The load balancer scans the index table from top till the first VM available is found or the index table is scanned fully. If the VM is found, the load data center communicates the request to the VM identified by the id. Further, the data center recognizes the load balancer of the new provision and the data center revises the index

table accordingly. When a request is allocated VM the current load on the VM is not taken in consideration which results in improvement of response time of a job.

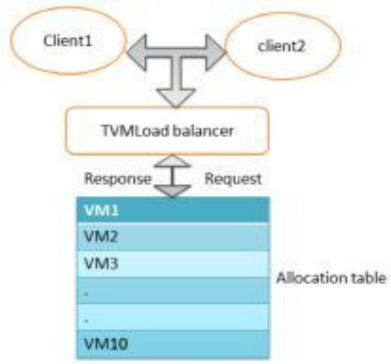

Figure 5: Throttled Algorithm

➢ **Randomized**

Randomized algorithm is static in nature. In this algorithm [16] a specific node n handles a process with a probability p. The order of process allocation for each processor is maintained independent of allocation from remote processor. This algorithm performs well where processes are equally loaded. Problem arises in the case when there are different computational complexities of loads. When load balancer receives a large number of requests, randomized algorithm can allocate the requests equally to the nodes Algorithm does not sustain deterministic approach. Like round robin algorithm, it is suitable for clusters containing nodes with similar configuration. It works well when Round Robin algorithm causes overhead for process queue. The approach followed in this algorithm is simpler, lightweight and very effective. It does not compute existing load on server so, it may cause selection of a VM which is heavily loaded and experience a long waiting time before being serviced [17]. No time management is required in this approach.

5. Metrics for load balancing algorithms

- Nature: This factor specifies the nature or behavior of load balancing algorithm that is whether the load balancing algorithm is of static or dynamic in nature.
- Throughput: It is the total amount of work done by all the nodes in a given time period. High throughput is required for better system performance.
- Fault Tolerance: An algorithm is said to be fault tolerant if it performs uniform load balancing even in case of any failure or breakdown. A good load balancing algorithm must be highly fault tolerant.
- Overload Rejection: It is the ability of an algorithm to reject incoming new requests in case of overload.
- Reliability: It is ability of an algorithm to continue its working perfectly in the case when one or more machine or node fails.
- Stability: It can be related with the delays in the transfer of information from one processor to other and the benefits of the load balancing algorithm which may help in getting a better performance.
- Execution System: An algorithm may be centralized or decentralized in nature. In centralized systems, global information is stored at a particular node. In distributed approach, every node performs load balancing individually.
- Response time: It is the time taken by a load balancing algorithm to give response. Or we can say that it is the amount of time between the request submission and arrival of response after completing the job. It must be low.
- Resource utilisation: It is a factor to check efficient utilisation of resources. Resource utilisation must be high to get better overall performance of the system.
- Complexity: Complexity describes the behaviour of a system. An algorithm is complex if it is difficult to understand and implement.

6. Comparison of different load balancing algorithms:

Table 1: Comparison of load balancing algorithms

Algorithms / Parameters	Round Robin	Throttled	ESCE	Randomized
Nature	Static	Dynamic	Dynamic	Static
Throughput	Low	Average	Average	Low
Fault Tolerant	No	No	No	No
Overload Rejection	No	Yes	No	No
Reliability	Less	Less	Less	Less
Stability	More	More	More	More
Execution System	Decentralized	Decentralized	Decentralized	Decentralized
Response time	More	Average	Average	Average
Resource utilisation	Less	High	High	Less
Complexity	Low	Low	Low	Minimum

The table compares different load balancing algorithms on the basis of metrics defined in previous section. Static load balancing techniques offer us easiest way to model and monitor cloud computing environment but it does not allow us to design heterogeneous nature of cloud. On the other side, dynamic load balancing techniques are hard to implement but are suitable for heterogeneous cloud environments. Round robin and randomized algorithms are static in nature and have lower throughputs whereas throttled and equal spread current execution load algorithms are dynamic in nature and have better throughputs. Only throttled algorithm has the property of overload rejection. Round robin algorithm serves the highest response time. Resource utilization is high in throttled and ESCE algorithms as compared to round robin and randomized. Randomized algorithm has lowest complexity in all of the four algorithms. After comparing algorithms with respect to different parameters we can say that static algorithms are more stable then dynamic ones but the later ones are better in performance. Here we can say that throttled load balancing algorithm is better on the basis of different metrics.

7. Conclusion

Cloud computing is emerging era in data management. It's very useful for managing the big data and shorting of data but due to load, performance may effect. That's why it is important to manage the load. Load balancing is very important for user satisfaction and resource utilization because it improves system performance that's why it becomes very important for research. To perform load balancing, various algorithms have been proposed i.e. round robin, ESCE etc. In this paper, various load balancing algorithms are discussed and compared on the basis of different metrics and we tried to find out the better load balancing algorithm.

References

1. Mell P, Grance T. The NIST definition of cloud computing.
2. Jadeja Y, Modi K. Cloud computing-concepts, architecture and challenges. InComputing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on 2012 Mar 21 (pp. 877-880). IEEE.
3. Wen YF, Chang CL. Load balancing consideration of both transmission and process responding time for multi-task assignment. International Journal of Space-Based and Situated Computing. 2014 Jan 1;4(2):100-13.
4. Alakeel AM. A guide to dynamic load balancing in distributed computer systems. International Journal of Computer Science and Information Security. 2010 Jun;10(6):153-60.
5. RajkumarBuyya, Christian Vecchiola, S. ThamaraiSelvi, Mastering Cloud Computing Foundations and Applications Programming. Elsevier, 2013.

6. RajkumarBuyya, James Broberg, and Andrzej M. Goseinski, Cloud Computing Principles and Paradigms. Wiley, 2011.
7. RajwinderKaur and PawanLuthra, "Load balancing in cloud computing," in Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing, ITC, 2012.
8. ShyamPatidar, DheerajRane and Pritesh Jain , "A Survey Paper on Cloud Computing," in Second International Conference on Advanced Computing & Communication Technologies(ACCT),2012
9. ShreyaPurohit, "An Exhaustive Study on Cloud Computing," volume II, ISSUE I, February 2016, ISSN NO. 2454-5678.
10. AnuRathi, Yogesh Kumar, AnishTalwar, "Aspects of Security in Cloud Computing," International Journal of Engineering and Computer Science, volume II, April 2013, ISSN 2278–0211.
11. Armbrust, Michael, et al. "A view of cloud computing." Communications of the ACM, volume 53, issue 4, april, 2010, pp. 50-58, ISSN: 0001-0782 EISSN: 1557-7317.
12. Asha M L, Neethu Myshri R, "Performance Evaluation of Round Robin Algorithm in Cloud Environment," in International Journal of Computer Applications (0975 – 8887) International Conference on Information and Communication Technologies (ICICT-2014).
13. Ray, Soumya, and Ajanta De Sarkar, "Execution analysis of load balancing algorithms in cloud computing environment," International Journal on Cloud Computing: Services and Architecture (IJCCSA) 2.5 (2012), pp.1-13.
14. Bhavisha Patel and Mr. Shreyas Patel, "Various Load Balancing Algorithms in cloud computing," IJARIIEISSN (O)-2395-4396, Vol-1 Issue-2 2015.
15. Khiyaita et al. "Load Balancing Cloud Computing: State Of Art," Network Security and Systems (JNS2), 2012 National Days of. IEEE, 2012.IEEE, 2010.
16. Xu, Zhong, and Rong Huang. "Performance study of load balancing algorithms in distributed web server systems." CS213 Parallel and Distributed Processing Project Report 1 (2009).
17. Isam Azawi Mohialdeen , "Comparative Study of Scheduling
Algorithms in Cloud Computing Environment ", Journal of Computer Science , 2013, pp. 252-263.
18. Amandeep Kaur Sidhu, Supriya Kinger, "Analysis of Load Balancing Techniques in Cloud Computing ", International Journal of Computers & Technology , Volume 4 No. 2, March-April, 2013, pp. 737-741.
19. M.Aruna , D. Bhanu, R.Punithagowri , "A Survey on Load Balancing Algorithms in Cloud Environment ", International Journal of Computer Applications, Volume 82 – No 16, November 2013 , pp. 39-43.
20. Padhy, Ram Prasad, and P. Rao. "Load balancing in cloud computing systems." PhD diss., 2011.
21. Behal, Veerawali, and Anil Kumar. "Cloud computing: Performance analysis of load balancing algorithms in cloud heterogeneous environment." Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-. IEEE, 2014.
22. Mohapatra, Subasish, K. Smruti Rekha, and Subhadarshini Mohanty. "A comparison of four popular heuristics for load balancing of virtual machines in cloud computing." International Journal of Computer Applications 68.6 (2013).
23. Wickremasinghe, Bhathiya, Rodrigo N. Calheiros, and Rajkumar Buyya. "Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications." Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on. IEEE, 2010.